The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP9 1RH

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed Andrew Gregg

Dated 16 APR 1999

Patents Form 1/77

Pat[redacted]977
(Rul[redacted]

The Patent Office

The Patent Office
Cardiff Road
Newport
Gwent NP9 1RH

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form))

1. Your reference — P74486 GCW PJC

2. Patent application number *(The Patent Office will fill in this part)* — **9806449.6** — 25 MAR 1998

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

PEPTIDE THERAPEUTICS LIMITED,
321 Cambridge Science Park,
Milton Road, Cambridge,
Cambridgeshire, CB4 4WG,
United Kingdom

Patents ADP number *(if you know it)* [redacted]

If the applicant is a corporate body, give the country/state of its incorporation

4. Title of the invention — ATTENUATED BACTERIA USEFUL IN VACCINES

5. Name of your agent *(if you have one)* — J A KEMP & CO

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

14 SOUTH SQUARE
GRAY'S INN
LONDON WC1R 5LX

Patents ADP number *(if you know it)*

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer "Yes" if:*
a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body:
See note (d))

Yes

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	44
Claim(s)	3
Abstract	1
Drawing(s)	6 + 6

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. I/We request the grant of a patent on the basis of this application

Signature A.BENTHAM for P.J. CAMPBELL, J.A. KEMP & Co Date 25 March 1998

12. Name and daytime telephone number of person to contact in the United Kingdom

P. J. CAMPBELL
0171 405 3292

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.

b) Write your answers in capital letters using black ink or you may type them.

c) If there is not enough space for all the relevant details on any part of this form, please continue of a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) If you have answered "Yes" Patents Form 7/77 will need to be filed.

e) Once you have filled in the form you must remember to sign and date it.

f) For details of the fee and ways to pay please contact the Patent Office.

ATTENUATED BACTERIA USEFUL IN VACCINES

The invention relates to attenuated bacteria useful in vaccines.

Background to the invention

The principle behind vaccination is to induce an immune response in the host thus providing protection against subsequent challenge with a pathogen. This may be achieved by inoculation with a live attenuated strain of the pathogen, i.e. a strain having reduced virulence such that it does not cause the disease caused by the virulent pathogen.

Clasically, live attenuated vaccine strains of bacteria and viruses have been selected using one of two different methodologies. Mutants have been created either by treatment of the organism using mutagenic chemical compounds or by repeated passage of the organism *in vitro*. However, use of either method gives rise to attenuated strains in which the mode of attenuation is unclear. These strains are particularly difficult to characterise in terms of possible reversion to the wild type strain as attenuation may reflect single (easily reversible) or multiple mutation events. Furthermore, it is difficult to obtain such strains having optimum immunogenic properties because of multiple mutation events, and multiple strains may need to be used to provide protection against the pathogen.

Using modern genetic techniques, it is now possible to construct genetially defined attenuated bacterial strains in which stable attenuating deletions can be created. A number of site directed mutants of Salmonella have been created using this type of technology (2, 4, 5, 9, 12, 16, 17, 18). Mutations in a large number of genes have

been reported to be attenuating, including the aro genes (e.g. *aroA, aroC, aroD* and *aroE*), *pur, htrA, ompR, ompF, ompC, galE, cya, crp* and *phoP*.

Salmonella *aroA* mutants have now been well characterised and have been shown to be excellent live vaccines against salmonellosis in several animal species. In addition, in order to reduce the chances of a reversion to virulence by a recombination event, mutations have been introduced into two independent genes such as *aroA/purA* and *aroA/aroC*. Identical mutations in host adapted strains of Salmonella such as *S.typhi* (man) and *S.dublin* (cattle) has also resulted in the creation of a number of candidate single dose vaccines which have proved successful in clinial (8, 11) and field trials (10).

A *Salmonella typhimurium* strain harboring stable mutations in both *ompC* and *ompF* is described in Chatfield *et al* (1991, ref. 21). When administered orally to BALB/c mice the strain was attenuated, with the 50% lethal dose (LD50) reduced by approximately 1,000-fold. However, the intravenous LD50 was reduced only by approximately 10-fold, demonstrating the importance of the porins in confering on the bacteria the ability to infect by the oral route.

Expression of the *ompC* and *ompF* genes is regulated by *ompR*. Pickard *et al* (1994, ref. 13) describes the cloning of the *ompB* operon, comprising the *ompR* and *envZ* genes, from a *Salmonella typhi* Ty2 cosmid bank and characterisation by DNA sequence analysis. The DNA sequence data were used to identify appropriate restriction sites for generating a defined deletion of 517 bp within the open reading frame of the *ompR* gene.

This deletion was introduced by homologous recombination into the chromosomes of two *S.typhi* strains which already harbored defined deletions in both the *aroC* and *aroD* genes. The *S.typhi ompR* mutants displayed a marked decrease in *ompC* and *ompF* porin expression as demonstrated by examination of outer membrane preparations. It was also shown that the *ompR-envZ* two component regulatory system plays an important role in the regulation of Vi polysaccharide synthesis in *S.typhi*.

In animal studies, attenuated *S.typhimurium* has been used as a vehicle for the delivery of heterologous antigens to the immune system (3, 6, 15). This raises the potential of the development of multivalent vaccines for use in man (7).

Summary of the Invention

The invention provides a bacterium attenuated by a non-reverting mutation in each of the *aroC* gene, the *ompF* gene and the *ompC* gene. The invention also provides a vaccine containing the bacterium.

It is believed that the *aroC/ompF/ompC* combination of mutations gives a vaccine having superior properties. For example, it is believed that the *aroC/ompF/ompC* combination may be superior to a *aroC/ompR* combination for two reasons:

1. The *ompR* mutation may cause higher levels of attenuation than the *ompF/ompC* combination of mutations because *ompR* may regulate a number of genes other than *ompF* and *ompC* which are important for survival of the bacterium *in vivo*. Thus, the *ompF/ompC* combination may allow the bacterium to

survive in the vaccinated host for a longer time and at higher levels, resulting in better protection.

2. The *ompR* mutation may cause reduced immunogenicity compared to the *ompF/ompC* combination of mutations because *ompR* may regulate the expression of antigens important for immunogenicity.

Detailed Description of the Invention

Bacteria useful in the Invention

The bacteria that are used to make the vaccines of the invention are generally those that infect by the oral route. The bacteria may be those that invade and grow within eukaryotic cells and/or colonise mucosal surfaces. The bacteria are generally Gram-negative.

The bacteria may be from the genera Escherichia, Salmonella, Vibrio, Haemophilus, Neisseria, Yersinia, Bordetella or Brucella. Examples of such bacteria are *Escherichia coli* - a cause of diarrhoea in humans; *Salmonella typhimurium* - the cause of salmonellosis in several animal species; *Salmonella typhi* - the cause of human typhoid; *Salmonella enteritidis* - a cause of food poisoning in humans; *Salmonella choleraesuis* - a cause of salmonellosis in pigs; *Salmonella dublin* - a cause of both a systemic and diarrhoel disease in cattle, especially of new-born calves; *Haemophilus influenza* - a cause of meningitis; *Neisseria gonorrhoeae* - a cause of gonorrhoeae; *Yersinia enterocolitica* - the cause of a spectrum of diseases in humans ranging from gastroenteritis to fatal septicemic disease; *Bordetella pertussis* - the cause of whooping cough; and *Brucella*

abortus - a cause of abortion and infertility in cattle and a condition known as undulant fever in humans.

Strains of *E.coli* and Salmonella are particularly useful in the invention. As well as being vaccines in their own right against infection by Salmonella, attenuated Salmonella can be used as carriers of heterologous antigens from other organisms to the immune system via the oral route. Salmonella are potent immunogens and are able to stimulate systemic and local cellular and antibody responses. Systems for driving expression of heterologous antigens in Salmonella *in vivo* are known; for example the *nirB* and *htrA* promoters are known to be effective drivers of antigen expression *in vivo*.

The invention may be appied to enterotoxigenic *E.coli* ("ETEC"). ETEC is a class of *E.coli* that cause diarrhoea. They colonise the proximal small intestine. A standard ETEC strain is ATCC H10407.

Infections of ETEC are the single most frequent cause of travellers diarrhoea, causing 3-9 million cases per year amongst visitors to developing countries. In endemic areas, ETEC infections are an important cause of dehydrating diarrhoea in infants and young children, resulting in 800,000 deaths a year in the under fives world-wide. In developing countries, the incidence of ETEC infections leading to clinical disease decreases with age, indicating that immunity to ETEC infection can be acquired. In contrast, naive adults from industrialized countries who visit endemic areas are highly susceptible to ETEC infections. However, with prolonged or repeated visits to endemic areas susceptibility to ETEC infections diminishes, suggesting that a live attenuated approach to ETEC vaccination may

prove successful.

The inventors chose to work on a non-toxigenic strain of ETEC called E1392/75/2A. E1392/75/2A arose spontaneously from a toxic mutant by deletion of toxin genes. In human studies, oral vaccination with live E1392/75/2A gave 75% protection against challenge with toxin-expressing ETEC from a different serotype. However, approximately 15% of vaccinees experienced diarrhoea as a side effect of the vaccine. The strain needs further attenuation to reduce the side effects before it can be considered as a potential vaccine and the invention gives a means of achieving such attenuation.

Seq Id No. 1 shows the sequence of the *E.coli aroC* gene, Seq Id No. 3 shows the sequence of the *E.coli ompC* gene and Seq. Id No. 5 shows the sequence of the *E.coli ompF* gene.

Further mutations

One or more further mutations may be introduced into the bacteria of the invention to generate strains containing mutations in addition to those in *aroC, ompC* and *ompF*. Such a further mutation may be (i) an attenuating mutation in a gene other than *aroC, ompC* and *ompF*, (ii) a mutation to provide *in vivo* selection for cells maintaining a plasmid (e.g. a plasmid expressing a heterologous antigen), or (iii) a mutation to prevent expression of a toxin gene.

The further attenuating mutation may be a mutation that is already known to be attenuating. Such mutations include mutations in *aro* genes (e.g. *aroA, aroD* and

aroE), pur, htrA, ompR, galE, cya, crp, phoP and *surA* (see e.g. refs 2, 4, 5, 9, 12, 13, 16, 17 and 18).

A mutation to provide selection for maintenance of a plasmid may be made by mutating a gene that is essential for the bacterium to survive. A plasmid carrying the essential gene is then introduced into the bacterium, so that only cells carrying the plasmid can survive. This may be useful where the plasmid contains, for example, a heterologous antigen to be expressed by the bacterium.

A mutation to prevent expression of a toxin gene may be made to reduce any side-effects caused by vaccination with the bacterium. For example, in the case of vaccination with *E.coli* strains such as ETEC it may be desirable to mutate the heat labile toxin (LT) or heat stable toxin (ST) genes so that they are not expressed.

The nature of the mutations

The mutations introduced into the bacterial vaccine generally knock-out the function of the gene completely. This may be achieved either by abolishing synthesis of any polypeptide at all from the gene or by making a mutation that results in synthesis of non-functional polypeptide. In order to abolish synthesis of polypeptide, either the entire gene or its 5'-end may be deleted. A deletion or insertion within the coding sequence of a gene may be used to create a gene that synthesises only non-functional polypeptide (e.g. polypeptide that contains only the N-terminal sequence of the wild-type protein).

The mutations are non-reverting mutations. These are mutations that show essentially no reversion back to the

wild-type when the bacterium is used as a vaccine. Such mutations include insertions and deletions. Insertions and deletions are preferably large, typically at least 10 nucleotides in length, for example from 10 to 600 nucleotides. Preferably, the whole coding sequence is deleted.

The bacterium used in the vaccine preferably contains only defined mutations, i.e. mutations which are characterised. It is clearly undesirable to use a bacterium which has uncharacterised mutations in its genome as a vaccine because there would be a risk that the uncharacterised mutations may confer properties on the bacterium that cause undesirable side-effects.

The attenuating mutations may be introduced by methods well known to those skilled in the art (see ref. 14). Appropriate methods include cloning the DNA sequence of the wild-type gene into a vector, e.g. a plasmid, and inserting a selectable marker into the cloned DNA sequence or deleting a part of the DNA sequence, resulting in its inactivation. A deletion may be introduced by, for example, cutting the DNA sequence using restriction enzymes that cut at two points in or just outside the coding sequence and ligating together the two ends in the remaining sequence. A plasmid carrying the inactivated DNA sequence can be transformed into the bacterium by known techniques such as electroporation and conjugation. It is then possible by suitable selection to identify a mutant wherein the inactivated DNA sequence has recombined into the chromosome of the bacterium and the wild-type DNA sequence has been rendered non-functional by homologous recombination.

Expression of heterologous antigens

The attenuated bacterium of the invention may be genetically engineered to express an antigen that is not expressed by the native bacterium (a "heterologous antigen"), so that the attenuated bacterium acts as a carrier of the heterologous antigen. The antigen may be from another organism, so that the vaccine provides protection against the other organism. A multivalent vaccine may be produced which not only provides immunity against the virulent parent of the attenuated bacterium but also provides immunity against the other organism. Furthermore, the attenuated bacterium may be engineered to express more than one heterologous antigen, in which case the heterologous antigens may be from the same or different organisms.

The heterologous antigen may be a complete protein or a part of a protein containing an epitope. The antigen may be from another bacterium, a virus, a yeast or a fungus. More especially, the antigenic sequence may be from *E.coli* (e.g. ETEC),tetanus, hepatitis A, B or C virus, human rhinovirus such as type 2 or type 14, herpes simplex virus, poliovirus type 2 or 3, foot-and-mouth disease virus, influenza virus, coxsackie virus or *Chlamydia trachomatis*. Useful antigens include non-toxic components of *E.coli* heat labile toxin, *E.coli* K88 antigens, ETEC colonization factor antigens, P.69 protein from *B.pertussis* and tetanus toxin fragment C.

The ETEC colonization factors and components thereof are prime candidates for expression as heterologous antigens. To instigate diarrhoeal disease, pathogenic strains of ETEC must be able to colonize the intestine and elaborate enterotoxins. For most strains of ETEC colonization factors (CF) that are required for adhesion to the

intestinal mucosa have been identified. In almost all cases CFs are expressed as fimbrae on the outer surface of the bacteria. A large number of CFs have been identified, the most prevalent being CFAI, CRAII (includes CS1, CS2, CS3) and CFAIV (includes CS4, CS5, CS6).

A vaccine to ETEC will ideally give protection against a range of colonization factor antigens to ensure that protection against different strains is obtained. In order to achieve this, it would be possible to express several colonization factors in one strain. Alternatively, the same attenuations could be made in a range of different ETEC strains, each with a different colonization factor. This would involve deleting the toxins from such strains.

The DNA encoding the heterologous antigen is expressed from a promoter that is active *in vivo*. Two promoters that have been shown to work well in Salmonella are the *nirB* promoter (19, 20) and the *htrA* promoter (20). For expression of the ETEC colonization factor antigens, the wild-type promoters could be used.

A DNA construct comprising the promoter operably linked to DNA encoding the heterologous antigen may be made and transformed into the attenuated bacterium using conventional techniques. Transformants containing the DNA construct may be selected, for example by screening for a selectable marker on the construct. Bacteria containing the construct may be grown *in vitro* before being formulated for administration to the host for vaccination purposes.

Formulation of the vaccine

The vaccine may be formulated using known techniques for formulating attenuated bacterial vaccines. The vaccine is advantageously presented for oral administration, for example in a lyophilised encapsulated form. Such capsules may be provided with an enteric coating comprising, for example, Eudragate "S" (Trade Mark), Eudragate "L" (Trade Mark), cellulose acetate, cellulose phthalate or hydroxypropylmethyl cellulose. These capsules may be used as such, or alternatively, the lyophilised material may be reconstituted prior to administration, e.g. as a suspension. Reconstitution is advantageously effected in a buffer at a suitable pH to ensure the viability of the bacteria. In order to protect the attenuated bacteria and the vaccine from gastric acidity, a sodium bicarbonate preparation is advantageously administered before each administration of the vaccine. Alternatively, the vaccine may be prepared for parenteral administration, intranasal administration or intramuscular administration.

The vaccine may be used in the vaccination of a mammalian host, particularly a human host but also an animal host. An infection caused by a microorganism, especially a pathogen, may therefore be prevented by administering an effective dose of a vaccine prepared according to the invention. The dosage employed will ultimately be at the discretion of the physician, but will be dependent on various factors including the size and weight of the host and the type of vaccine formulated. However, a dosage comprising the oral administration of from 10^7 to 10^{11} bacteria per dose may be convenient for a 70 kg adult human host.

Experimental section

The experiments described in this section serve to

illustrate the invention.

Brief description of the drawings

Figure 1 shows a system for constructing defined deletions in target genes using splicing by overlay extension PCR mutagenesis.

Figure 2 shows the expected sequences of target genes after recombination and selection for deletions.

Figure 3 shows the cloning of deletion cassettes into plasmid pCVD442.

Figure 4 shows an SDS-PAGE analysis of outer membranes prepared from ETEC strains under conditions of low (no salt L-broth) and high (no salt L-broth + 15% sucrose) osmolarity. M = markers; Sample 1 = PTL010; Sample 2 = PTL002; Sample 3 = PTL003; Sample 4 = ΔaroCΔompC; Sample 5 = ΔompF.

Figure 5 shows expression of CS1 and CS3 in deletion strains after growth on CFA agar. Equal numbers of cells from each strain were loaded on a 15% SDS-PAGE gel and Western blotted with monospecific anti-CS1 or anti-CS3 polycional antibodies. Controls for antibody specificity were whole cesll lysates of TG1 cells expressing the majore pilin protein of CS1, or purified major pilin protein from CS3. Lane M, rainbow low molecular mass markers; lane 1, induced TG1 cells harbouring pKK223; lane 2, induced TG1 cells harbouring pKKCs1; lane 3, CS1-ETEC strain; lane 4, PTL010; lane 5, PTL001; lane 6, PTL002; lane 7, PTL003; lane 8, purified CS3 major pilin protein.

Figure 6 shows a Southern blot of mutant loci. Chromosomal DNA was extracted from the wild-type ETEC (E1392/75-2A), PTL001 (htrA aroC), PTL002 (aroC ompR) and PTL003 (aroC ompC ompF) as indicated, digested with restriction endonuclease EcoRV, and pulsed field electrophoresed through 1% agarose. DNA was blotted from the gel onto Hybond N+ nylon membranes (Amersham) and hybridised with DNA probes derived from the aroC, htrA, ompR, ompC, or ompF loci as shown. The banding patterns are consistent with the mutant loci being deletions.

Design of deletions and construction of plasmids pCVDΔ*AroC*, pCVDΔ*OmpC* and pCVDΔ*OmpF*

Deletions were designated to remove the entire open reading frame of the target gene. Using the *E.coli* genome sequence as a template, PCR primers were designed to amplify fragments of 500-600 base pairs flanking the target open reading frame (see Table 1 for primer sequences). Splicing by overlap extension using PCR was used to fuse the two flanking sequences, creating a PCR product with the entire gene deleted (Figure - 1). The wild-type sequences around the deletion site and the predicted sequences after deletion are depicted in Figure 2.

For each gene two different restriction sites were introduced into the splice region (see Table 2 below). These were used for identification of deletion clones. The PCR primers at either end of the PCR fragment introduced unique restriction sites that were used to clone the fragment into the multiple cloning site of pCVD442 (Figure - 3).

PCR products were gel purified using a Qiagen (Trade Name) gel extraction kit and digested with the relevant

restriction enzymes prior to ligation to the suicide plasmid pCVD442(22) digested with the same enzyme and treated with alkaline phosphatase to prevent vector self-ligation (Figure - 3). The ligation mix was transformed into SY327λ*pir* and plated on L-Ampicillin (100 µg/ml) plates. Plasmids from Ampicillin resistant transformants were screened for the presence of the deletion cassettes by restriction digestion. The following plasmids were generated:

pCVD*ΔAroC*
pCVD*ΔOmpC*
pCVD*ΔOmpF*

The suicide plasmid pCVD442 can only replicate in cells harboring the *pir* gene. On introduction into *non-pir* strains, pCVD442 is unable to replicate, and the Ampicillin resistance conferred by the plasmid can only be maintained if the plasmid is integrated in the chromosome by a single homologous recombination event. The plasmid also has a *sacB* gene, encoding levan sucrase, which is toxic to gram negative bacteria in the presence of sucrose. This can be used to select clones that have undergone a second recombination event, in which the suicide plasmid is excised. Such cells will be resistant to sucrose, but Ampicillin sensitive.

Construction and characterisation of *ΔAroCΔOmpCΔOmpF* strain

This section outlines the chronology of construction and history of a *ΔAroCΔOmpCΔOmpF* strain. In the section, "ETEC" refers specifically to strain E1392/75/2A or its derivatives.

ΔAroCΔOmpCΔOmpF deletions were introduced into E1392/75/2A in the following order: *ΔAroC→ΔAroCΔOmpC→ΔAroCΔOmpCΔOmpF*

Construction of ETEC*ΔAroC*

1) E1392/75/2A from original microbanked stock was plated onto L-Agar.
2) Electroporation competent cells were prepared from these cells. 100 µl aliquots were frozen.
3) pCVD*ΔAroC* was purified from SY327*pir* cells using a Qiagen Qiafilter (Trade Name) midiprep. The plasmid was concentrated about 10-fold by ethanol precipitation. The construction of pCVD*ΔAroC* is described above.
4) 5 µl of concentrated plasmid was mixed with 100 µl defrosted cells and electroporated. The whole transformation was plated on an L-Ampicillin plate (50 µg/ml) and incubated overnight at 37°C.
5) A single Ampicillin resistant colony grew.
6) The colony was streaked onto an L-Ampicillin plate (100 µg/ml) and grown overnight at 37°C ("merodiploid plate").
7) PCR using primers TT19 and TT20 (specific for the *aroC* gene) and a colony picked from the merodiploid plate amplified two bands, with sizes corresponding to that of the wild-type and *ΔaroC* genes. The sequences of the primers are shown in Table 1 below.
8) A colony from the merodiploid plate was grown up for 7 hr in a) L-Ampicillin broth (100 µg/ml) and b) L-Broth. The colony grown on L-Ampicillin was microbanked.
9) Serial dilutions of the L-broth culture were set up on:
    a) No salt L-agar

b) No salt L-agar + 5% sucrose.

The plates were incubated overnight at 30°C.

10) Colony counts showed that 10^4 more colonies grew on L-agar than on L-agar + 5% sucrose, showing sucrose selection worked.

11) Sucrose resistant colonies were screened for the presence of *ΔaroC* gene by PCR. Colonies chosen for screening were picked onto an L-agar plate and grown overnight at 37°C. This plate was stored at 4°C, whilst further tests were carried out.

12) 50% of 90 colonies tested had *ΔaroC* only.

13) Colonies were tested for growth on:

a) M-9 minimal media plates

b) M-9 minimal media + Aromix plates

c) L-Amp (100 µg/ml)

ΔaroC colonies should not grow on M-9 minimal media without Aromix or on L-Amp.

Aromix is a mix of aromatic compounds as follows:

Substance	Final concentration (% w/v)
Phenylalanine	0.004
Tryptophan	0.004
Tyrosine	0.004
p-aminobenzoic acid	0.001
dihydroxybenzoic acid	0.001

These compounds are made in wild-type bacteria, but the *aroC* mutation prevents their synthesis.

14) 13/14 putative ΔAroC colonies required Aromix for growth on M-9 minimal media and were susceptible to Ampicillin.

15) 3 colonies (No. 1,2,3) were tested for the presence of the CS1 major pilin protein gene by PCR using

primers MGR169 and MGR170. All 3 colonies gave PCR products of the expected size (700 bp.). The sequences of the primers are shown in Table 1.

16) Colonies 1, 2 and 3 from screening master plate were streaked onto L-Agar and grown overnight at 37°C. Cells from these plates were used to inoculate microbank tubes.

17) Colony 1, stored in a microbank, was used for further work.

18) For permanent storage, a bead from the microbank tray was inoculated into 1 ml L-broth, grown for 4 hr with shaking at 37°C and used to make agar slopes which were used to make freeze dried stocks. The freeze dried stock of E1392/75/2A*ΔAroC* was designated PTL004. 20 ml of L-broth was added to the rest of the 1 ml culture and the culture was incubated overnight at 30°C. 1 ml of the overnight culture was transferred to each of three cryovials and stored in liquid nitrogen.

Construction of ETEC*ΔAroCΔOmpr*

1) Preparation of pCVD*ΔOmpC* plasmid DNA for electroporation:

A colony of SY327λ*pir* harbouring pCVD*ΔOmpC* was grown overnight at 37°C in 100 ml L-Ampicillin broth
(100 µg/ml). Plasmid DNA was purified using 2 Qiagen Qiafilter (Trade Name) midipreps. DNA was further concentrated by ethanol precipitation. The construction of pCVD*ΔOmpC* is described above.

2) Preparation of electrocompetent cells:
ETEC*ΔAroC* cells from the microbank tray produced in step 17 of the preceding section were streaked on L-agar, grown at 37°C overnight and then stored at 4°C for no more than 1 week before being used to

inoculate cultures for preparing electrocompetent cells.

3) ETEC*ΔAroC* cells were electroporated with 5 µl of concentrated pCVD*ΔOmpC* DNA, and each transformation plated on a single L-Ampicillin plate (50 µg/ml) and grown overnight at 37°C.

4) 17 Ampicillin resistant colonies (putative ETEC*ΔAroC*/ pCVD*ΔOmpC* merodiploids) were obtained.

5) These colonies were spotted onto a master L-Ampicillin (100 µg/ml) plate and used as templates for PCR with primers TT7/TT8. The master plate was grown at room temperature over the weekend. The sequences of the primers are given in Table 1 below.

6) A single colony (No. 7) had the *ΔompC* gene.

7) The colony was grown for 5 hr in L-broth.

8) Serial dilutions of the L-broth culture were set up on:

   a) No salt L-agar

   b) No salt L-agar + 5% sucrose.

The plates were incubated overnight at 30°C.

9) Colony counts showed that 10^4 more colonies grew on L-agar than on L-agar + 5% sucrose, showing sucrose selection worked.

10) 45 sucrose resistant colonies were screened for *ΔompC* by PCR using primers TT7 and TT8. 9 colonies had the *ΔompC* gene, but most had traces of w.t. *ompC gene*. The sequences of the primers are given in Table 1 below.

11) To further characterise putative ETECΔAroCΔOmpC colonies, they were grown in 1 ml L-Broth for 5 hr and plated on:

   a) L-Agar + 100 µg/ml Ampicillin

   b) L-Agar

   c) L-Agar + 5% sucrose

ΔOmpC colonies should be resistant to sucrose and sensitive to Ampicillin.

12) Only 1 colony (No. 1) was Ampicillin sensitive and sucrose resistant.

13) Colony 1 was checked for the presence of *ΔaroC, ΔompC* and CS1 genes by PCR with primers TT19/TT20, TT7/TT8 and MGR169 and 170. The sequences of the primers are given in Table 1 below.

14) Colony 1 gave single PCR products of the expected size for *ΔaroC, ΔompC* and CS1 genes.

15) The colony was microbanked.

16) For permanent storage, a bead from the microbank was inoculated into 1 ml L-broth, grown for 4 hr with shaking at 37°C and used to make agar slopes which were freeze dried. The freeze dried stock of E1392/75/2A*ΔAroCΔOmpC* was designated PTL008. 20 ml of L-broth was added to the rest of the 1 ml culture and the culture was incubated overnight at 30°C. 1 ml of the overnight culture was transferred to each of three cryovials and stored in liquid nitrogen.

Construction of ETEC*ΔAroCΔOmpCΔOmpF*

Conjugation was used to introduce pCVD*ΔOmpF* into E1392/75/2A*ΔAroCΔOmpC*.

1) Conjugation donor cells SM10λ*pir* were transformed with pCVD*ΔOmpF*. The construction of plasmid pCVD*ΔOmpF* is described above.

2) ETEC*ΔAroCΔOmpC* cells were conjugated with SM10λ*pir*/ pCVD*ΔOmpF* cells. The pCVD442 plasmid includes a transfer origin which allows the plasmid to be transferred from a donor strain containing the RP4 transfer genes (e.g. SM10λpir) to a recipient strain (e.g. ETEC). ETEC*ΔaroCΔompC* cells and

E.coli strain SM10λ*pir* harbouring the PcvdΔompF recombinant were cross-streaked on L-agar plates so as to cover an area of approximately 10 cm^2. Plates were incubated at 37° C for 20 h, then the growth washed off using 4 ml L-broth and the suspension plated onto McConkey agar (Difco) containing streptomycin at 20μg ml^{-1} and ampicillin at 300μg ml^{-1}. Plates were incubated overnight at 37°C and resulting colonies were checked for merodiploidy by PCR using appropriate oligonucleotides as primers.

3) Putative ETEC transconjugants were screened. 10 colonies were picked from McConkey agar plates and grown overnight on L-Ampicillin (100 µg/ml) agar. The presence of *ΔompF* gene was checked for by PCR with primers TT1/TT2. The sequences of the primers are given in Table 1 below.

4) The colonies were grown for 5 hr in L-broth.

5) Serial dilutions of the L-broth culture were set up on:

   a) No salt L-agar

   b) No salt L-agar + 5% sucrose.

The plates were incubated overnight at 30°C.

6) Colony counts showed 10^5 more colonies grew on L-agar than on L-agar + 5% sucrose, showing sucrose selection worked.

7) Sucrose resistant colonies were screened for *ΔompF* gene by PCR with primers TT1/TT2. The sequences of the primers are given in Table 1 below. The screened colonies were grown overnight on L-Agar. 3 colonies out of 47 had the *ΔompF* gene with no evidence of the wild-type *ompF* gene.

8) To further characterise putative ETEC*ΔAroCΔOmpCΔOmpF* colonies, they were plated on:

   a) L-Agar + 100 µg/ml Ampicillin

b) L-Agar

c) L-Agar + 5% sucrose

ΔompF colonies should be resistant to sucrose and sensitive to Ampicillin.

9) All three *ΔompF* colonies were Ampicillin sensitive and sucrose resistant.

10) The colonies were microbanked and one colony was chosen as a master stock.

11) For permanent storage, a bead from the master stock was inoculated into 1 ml L-broth, grown for 4 hr with shaking at 37°C and used to make agar slopes which were used to make freeze dried stocks. The freeze dried stock of E1392/75/ 2A*ΔaroCΔompCΔompF* was designated PTL003. 20 ml of L-broth was added to the rest of the 1 ml culture and the culture was incubated overnight at 30°C. 1 ml of the overnight culture was transferred to each of three cryovials and stored in liquid nitrogen.

Characterisation of E1392/75/2A*ΔAroCΔOmpCΔOmpF*

1) Growth requirements:

Cells taken from the master stock produced in step 10 of the preceding section were streaked on L-Agar plate. At the same time 8 ml L-broth was inoculated for a chromosomal DNA prep for Southern blots. Both plate and liquid culture were grown overnight at 37°C.

Cells from the grown plate were streaked onto the following media and grown overnight at 37°C.

Medium	Growth
L-Amp	No
M9 minimal media	No

M9 minimal + Aromix

Yes

M9 + sulfathiazole (100 µg/ml)

No

M9 + sulfathiazole (100 µg/ml) + Aromix

Yes

L-Agar + 50 µg/ml streptomycin

Yes

L-Agar + 5% sucrose

Yes

As expected, the cells were Amp sensitive. The cells were resistant to sucrose, streptomycin and sulfathiazole, but required Aromix to grow on minimal media.

2) LPS analysis of PTL003:

a) A freeze dried vial of PTL003 was broken open. The culture was resuspended in L-Broth and plated on L-Agar for growth. Some cells were scraped off and stored in microbank.

b) More cells were scraped off and the LPS profile was analysed. There was no visible difference between the LPS profile of PTL003 and original E1392/75/2A strain.

3) Confirmation of deletions by PCR:

a) A scrape of cells was taken from the plate made in in 2a and streaked onto L-Agar and grown overnight.

b) Freshly grown cells were used for PCR with primers that flank the following genes: *aroC, htrA, ompC, ompF, ompR*.

c) PTL003 was shown to have deletions in *aroC, ompC* and *ompF* genes, but not in *htrA* or *ompR*.

4) Outer membrane protein preps of PTL003:

a) An outer membrane prep of PTL003 was made

using cells from step 3a above.

b) The outer membrane fraction was frozen and gels run (Figure 4).

5) Checking expression of CS1 and CS3 in PTL003:

a) The plate from 3a above was used. A colony of PTL003 and PTL010 (E1392/75/2A freeze dried stock) were grown up for 4 hr in L-Broth. 2 µl was dotted on each of four CFA-Agar plates and grown overnight at 37°C or 18°C.

b) The 37°C and 18°C plates were blotted with anti-CS1 and anti-CS3 antibodies.

c) The results are shown in Figure 5. No CS1 or CS3 expression was seen at 18°C with either PTL010 or PTL003. Both PTL010 and PTL003 expressed CS1 and CS3 at 37°C. If anything PTL003 expressed slightly more CS1, but this may reflect different cell numbers or stickiness to nitrocellulose rather than differences in pili/cell.

6) Southern blotting of PTL003:

Structure of deletion mutations. Total DNA was extracted from cultures of the three deletion mutants grown from the microbanked stocks, digested with restriction endonuclease *Eco*RV, and the digested DNA subjected to pulsed field agarose gel electrophoresis. DNA was blotted from the gels onto Hybond N+ (Trade Name) nylon membranes and hybridised with appropriate DNA probes according to standard procedures. Results (Figure 6) show that the hybridising chromosomal DNA fragments of the mutants are shorter than the wild-type, consistent with the mutations being deletions. *Confirmation of absence of Heat-Stable (ST) and Heat-Labile (LT) toxin genes in E.coli strain E1392/75-2A.* For this the ST and LT-AB genes were used as DNA probes against total DNA from E1392/75-2A. Total DNA from the toxin positive ETEC

strain E1393/75 was included as a positive control, while that from the laboratory *E.coli* strain JM109 was included as a negative. Hybridised membranes were left under Hyperfilm-ECL (Trade Name) for 1 h to obtain the maximum amount of signal. Probes were prepared using PCR with plasmid DNA extracted from E1392/75-2A as template and oligonucleotides EST01 and EST02 as primers for ST, or LT-R1 and LT-03 for LT-AB. There was no significant hybridisation with total DNA using either the LT-AB or the ST probe, despite obtaining a very intense signal from the positive control total DNA.

Confirmation of absence of pCVD442 sequences from the chromosome of deletion mutants. The plasmid pCVD442 was labelled and hybridised to total DNA from deletion mutants PTL001, PTL002 and PTL003 digested with *Eco*RV. Total DNA from ETEC strain E1392/75-2A was included as a control. A complex pattern of hybridising DNA fragments was obtained. But, there was no significant difference between the pattern obtained for the wild-type and that for the mutants, indicating that probably no residual pCVD442 nucleotide sequences were left in the genomes of the mutants. The complex pattern of hybridising fragments was most likely due to the pCVD442 probe hybridising with the plasmid DNA components of the E1392/75-2A strain and mutant derivatives.

Analysis of outer membrane protein profile of PTL003: Outer membrane protein fractions were prepared from strains PTL010 (E1392/75/2A) and the deletion strains PTL002 and PTL003. A strain with a single *ompF* deletion and a strain with both *aroC* and *ompC* deletion were also analysed. Strains were grown under conditions of low osmolarity (no salt L-broth) and high osmolarity (no salt L-broth+15% sucrose). The OmpF protein product is normally expressed at low osmolarity whereas the OmpC

product is expressed at high osmolarity. The OmpC and OmpF proteins have similar electroporetic mobilities. At both high and low osmolarities, the strain PTL003 lacks proteins in the OmpC/OmpF region when compared to the wild-type E1392/75/2A strain or to the ΔAroCΔOmpC or ΔOmpF deletion strains. The results are shown in Figure 4.

Expression of CS1 and CS3 pili on CFA agar:

The expression of CS1 and CS3 pili in the deletion strains was examined. Equal numbers (2 A_{600nm} units) of bacteria strains PTL010, PTL001, PTL002 and PTL003 grown overnight at 37°C on CFA agar were subjected to SDS PAGE and analysed by Western blotting with monospecific polyclonal antibodies against CS1 or CS3. CS1 and CS3 pili were expressed equally well in four strains (Figure 5).

A CFAII-negative derivative of E1392/75/2A was constructed for use as a control. This was done by specific curing of the CS encoding plasmids from ETEC strain E1392/75-2A. A short fragment of DNA was amplified from the *cooB* gene using PCR with oligonucleotides CSA01 and CSA02 as primers and ligated into pGEM-T Easy plasmid vector (Trade Name, Promega) designed for the cloning of PCR products. The fragment was subcloned into pCVD442 by virtue of the *Sal*I and *Sph*I restriction enzyme sites. The pCVD442-*cooB* derivative was introduced into ETEC strain E1392/75/2A by conjugation from SM10λ*pir*. Ampicillin resistant transconjugants are most likely to be the result of fusion of the pCVD442-*cooB* derivative with *cooB*-bearing plasmid. Such transconjugates were then grown on L-agar supplemented with 5% sucrose to select for loss of the sacB gene of pCVD442. Resulting colonies were tested for ampicillin sensitivity, and by PCR using CSA01 and CSA02 as primers. Three colonies of E1392/75/2A

were included as positive controls among these PCRs. Two sucrose resistant colonies that gave no product with the PCR were streaked out onto fresh L-agar supplemented with 5% sucrose to obtain pure cultures. These were then grown in L-broth at 37°C for approximately 16 h and microbanked at -70°C. Loss of the CS1 encoding plasmid was confirmed by analysis of the plasmid profiles of the derivatives using agarose gel electrophoresis. Two derivatives were confirmed as CS1 negative, but were still CS3+.

Table 1 - **PCR primers**

Name	Target	Use	Sequence (5'-3')
TT1	*ompF*	Primer A for cloning	ATC TGT TTG TTG AGC TCA GCA ATC TAT TTG CAA CC
TT2	*ompF*	Primer B for cloning	TTT TTT GCC AGC ATG CCG GCA GCC ACG CGT AGT G
TT3	*ompF*	Primer C for cloning	CTC GAG GCT TAG CTC TAT TTA TTA CCC TCA TGG
TT4	*ompF*	Primer D for cloning	GAG CTA AGC CTC GAG TAA TAG CAC ACC TCT TTG
TT7	*ompC*	Primer A for cloning	TTG CTG GAA AGT CGA CGG ATG TTA ATT ATT TGT G
TT8	*ompC*	Primer B for cloning	GGC CAA AGC CGA GCT CAT TCA CCA GCG GCC CGA CG
TT9	*ompC*	Primer C for cloning	GCT AAG CCT CGA GTA ATC TCG ATT GAT ATC CG
TT10	*ompC*	Primer D for cloning	CTC GAG GCT TAG CGT TAT TAA CCC TCT GTT A
TT19	*aroC*	Primer A for cloning	CCG CGC TCG CTC TAG AGT GAA CTG ATC AAC AAT A
TT20	*aroC*	Primer B for cloning	ATG CGC GCG AGA GCT CAA CCA GCG TCG CAC TTT G

TT21	*aroC*	Primer C for cloning	CTC GAG GCA TGC TGA ATA AAA CCG CGA TTG
TT22	*aroC*	Primer D for cloning	GCA TGC CCT CGA GGG CTCC GTT ATT GTT GTG
MGR169	CS1	Binds in CS1 sequence	TGA TTC CCT TTG TTG CGA AGG CGA A
MGR170	CS1	Binds in CS1 sequence	ATT AAG ATA CCC AAG TAA TAC TCA A
LT-R1	LT-AB	See text	GCT TTT AAA GGA TCC TAG TT
LT-03	LT-AB	See text	GGT TAT CTT TCC GGA TTG TC
EST01	ST	See text	CAT GTT CCG GAG GTA ATA TGA A
EST02	ST	See text	AGT TCC CTT TAT ATT ATT AAT A
CSA01	CS1	See text	TGG AGT TTA TAT GAA ACT AA
CSA02	CS1	See text	TGA CTT AGT CAG GAT AAT TG
CS3-01	CS3	See text	ATA CTT ATT AAT AGG TCT TT
CS3-02	CS3	See text	TTG TCG AAG TAA TTG TTA TA

Table 2

Target gene	Sites used for cloning into pCVD442		Sites introduced for screening purposes	
	Site 1	Site 2	Site 3	Site 4
aroC	XbaI	SacI	XhoI	SphI
htrA	SalI	SphI	XhoI	XbaI
ompC	SalI	SacI	BlpI	XhoI
ompF	SacI	SphI	BlpI	XhoI
ompR	SalI	SacI	BlpI	SphI

References

1. Bacon, G.A., Burrows, T.W. and Yates, M. (1950) Br.J.Exp. Pathol., 31, 714-24

2. Chatfield, S.N., Charles, I.G., Makoff, A.J. et al (1992a) Biotech. 10, 888-892

3. Chatfield, S.N., Strahan, K., Pickard, D., Charles, I.G., Hormaeche, C.E. and Dougan, G. (1992b) Microbiol. Pathog., 12, 145-151

4. Curtiss III, R. and Kelly, S.M. (1987) Infect.Immun. 55, 3035-3043

5. Dougan, G., Chatfield, S., Pickard, D., Bester, J., O'Callaghan, D. and Maskell, D. (1988) J.Inf.Dis. 158, 1329-1335

6. Fairweather, N.F., Chatfield, S.N., Makoff, A.J. et al (1990) Infect.Immun. 58, 1323-1329

7. Gomaz-Duarte, O.G., Galen, J., Chatfield, S.N. (1995) Vaccine, 13:1596-1602

8. Hohmann, E.L., Oletta, C.A., Killeen, K.P. and Miller, S.I. (1996) Vaccine 14, 19-24

9. Hone, D., Morona, R., Attridge, S. and Hackett, J. (1987) J.Infect.Dis., 156, 167-1

10. Jones, P.W., Dougan, G., Haywood, C., MacKensie, N., Collins, P. and Chatfield, S.N. (1991) Vaccine 9, 29-36

11. Levine, M.M., Galen, J., Barry, E. et al (1995) J.Biotech, 44, 193-196

12. Miller, S.I., Kukral, A.M. and Mekalanos, J.J. (1989), Proc.Nat.Acad.Sci., USA 86, 5054-5058

13. Pickard, D., Li, J.L., Roberts, M., Maskell, D., Hone, D., Levine, M., Dougas, G. and Chatfield, S. (1994) Infection and Immunity 62, 3984-3993

14. Sambrook, J., Fritsch, E.F. and Maniatis, T., (1989)
Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

15. Strugnell, R.A., Dougan, G., Chatfield, S.N. et al (1992) Infect.Immun., 60, 3994-4002

16. EP-B-0322237 (Dougan et al)

17. EP-B-0400958 (Dougan et al)

18. EP-B-0524205 (Dougan et al)

19. WO 92/15689 (Charles et al)

20. Everest, P., Allen, J., Papakonstantinopoulou, A., Mastroeni, P., Roberts, M. and Dougan, G. (1995) FEMS Microbiol. Letts., 126, 97-101

21. Chatfield, S.N., Dorman, C.J., Hayward, C. and Dougan, G. (1991) Infection & Immunity 59, 449-452

22. Donnenberg, M.S. and Kaper, J.B. (1991) Infection and Immunity 59, 4310-4317

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: PEPTIDE THERAPEUTICS LIMITED
(B) STREET: 321 Cambridge Science Park, Milton Road
(C) CITY: Cambridge
(D) STATE: Cambridgeshire
(E) COUNTRY: United Kingdom
(F) POSTAL CODE (ZIP): CB4 4WG

(ii) TITLE OF INVENTION: ATTENUATED BACTERIA USEFUL IN VACCINES

(iii) NUMBER OF SEQUENCES: 6

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(v) CURRENT APPLICATION DATA:
APPLICATION NUMBER:

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1690 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: aroC of E.coli

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:492..1562

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GTCGACGCGG TGGATATCTC TCCAGACGCG CTGGCGGTTG CTGAACAGAA CATCGAAGAA 60

CACGGTCTGA TCCACAACGT CATTCCGATT CGTTCCGATC TGTTCCGCGA CTTGCCGAAA 120

GTGCAGTACG ACCTGATTGT CACTAACCCG CCGTATGTCG ATGCGAAGAT ATGTCCGACC 180

TGCCAAACAA TACCGCCACG AGCCGGAACT GGGCCTGGCA TCTGGCACTG ACGGCCTGAA 240

ACTGACGCGT CGCATTCTCG GTAACGCGGC AGATTACCTT GCTGATGATG GCGTGTTGAT 300

```
TTGTGAAGTC GGCAACAGCA TGGTACATCT TATGGAACAA TATCCGGATG TTCCGTTCAC     360

CTGGCTGGAG TTTGATAACG GCGGCGATGG TGTGTTTATG CTCACCAAAG AGCAGCTTAT     420

TGCCGCACGA GAACATTTCG CGATTTATAA AGATTAAGTA AACACGCAAA CACAACAATA     480

ACGGAGCCGT G ATG GCT GGA AAC ACA ATT GGA CAA CTC TTT CGC GTA ACC      530
             Met Ala Gly Asn Thr Ile Gly Gln Leu Phe Arg Val Thr
               1               5                  10

ACC TTC GGC GAA TCG CAC GGG CTG GCG CTC GGC TGC ATC GTC GAT GGT      578
Thr Phe Gly Glu Ser His Gly Leu Ala Leu Gly Cys Ile Val Asp Gly
     15                  20                  25

GTT CCG CCA GGC ATT CCG CTG ACG GAA GCG GAC CTG CAA CAT GAC CTC      626
Val Pro Pro Gly Ile Pro Leu Thr Glu Ala Asp Leu Gln His Asp Leu
 30                  35                  40                  45

GAC CGT CGT CGC CCT GGG ACA TCG CGC TAT ACC ACC CAG CGC CGC GAG      674
Asp Arg Arg Arg Pro Gly Thr Ser Arg Tyr Thr Thr Gln Arg Arg Glu
                 50                  55                  60

CCG GAT CAG GTC AAA ATT CTC TCC GGT GTT TTT GAA GGC GTT ACT ACC      722
Pro Asp Gln Val Lys Ile Leu Ser Gly Val Phe Glu Gly Val Thr Thr
             65                  70                  75

GGC ACC AGC ATT GGC TTG TTG ATC GAA AAC ACT GAC CAG CGC TCT CAG      770
Gly Thr Ser Ile Gly Leu Leu Ile Glu Asn Thr Asp Gln Arg Ser Gln
         80                  85                  90

GAT TAC AGT GCG ATT AAG GAC GTT TTC CGT CCA GGC CAT GCC GAT TAC      818
Asp Tyr Ser Ala Ile Lys Asp Val Phe Arg Pro Gly His Ala Asp Tyr
     95                  100                 105

ACC TAC GAA CAA AAA TAC GGT CTG CGC GAT TAT CGC GGC GGT GGA CGT      866
Thr Tyr Glu Gln Lys Tyr Gly Leu Arg Asp Tyr Arg Gly Gly Gly Arg
110                 115                 120                 125

TCT TCC GCC CGC GAA ACC GCC ATG CGC GTG GCG GCA GGA GCT ATT GCC      914
Ser Ser Ala Arg Glu Thr Ala Met Arg Val Ala Ala Gly Ala Ile Ala
                130                 135                 140

AAA AAA TAT CTC GCC GAG AAA TTT GGT ATT GAA ATC CGT GGC TGC CTG      962
Lys Lys Tyr Leu Ala Glu Lys Phe Gly Ile Glu Ile Arg Gly Cys Leu
            145                 150                 155

ACC CAG ATG GGC GAC ATT CCG CTG GAT ATC AAA GAC TGG TCG CAG GTC     1010
Thr Gln Met Gly Asp Ile Pro Leu Asp Ile Lys Asp Trp Ser Gln Val
        160                 165                 170

GAG CAA AAT CCG TTT TTT TGC CCG GAC CCC GAC AAA ATC GAC GCG TTA     1058
Glu Gln Asn Pro Phe Phe Cys Pro Asp Pro Asp Lys Ile Asp Ala Leu
    175                 180                 185

GAC GAG TTG ATG CGT GCG CTG AAA AAA GAG GGC GAC TCC ATC GGC GCT     1106
```

```
Asp Glu Leu Met Arg Ala Leu Lys Lys Glu Gly Asp Ser Ile Gly Ala
190                 195                 200                 205

AAA GTC ACC GTT GTT GCC AGT GGC GTT CCT GCC GGA CTT GGC GAG CCG      1154
Lys Val Thr Val Val Ala Ser Gly Val Pro Ala Gly Leu Gly Glu Pro
                210                 215                 220

GTC TTT GAC CGC CTG GAT GCT GAC ATC GCC CAT GCG CTG ATG AGC ATC      1202
Val Phe Asp Arg Leu Asp Ala Asp Ile Ala His Ala Leu Met Ser Ile
            225                 230                 235

AAC GCG GTG AAA GGC GTG GAA ATT GGC GAC GGC TTT GAC GTG GTG GCG      1250
Asn Ala Val Lys Gly Val Glu Ile Gly Asp Gly Phe Asp Val Val Ala
        240                 245                 250

CTG CGC GGC AGC CAG AAC CGC GAT GAA ATC ACC AAA GAC GGT TTC CAG      1298
Leu Arg Gly Ser Gln Asn Arg Asp Glu Ile Thr Lys Asp Gly Phe Gln
    255                 260                 265

AGC AAC CAT GCG GGC GGC ATT CTC GGC GGT ATC AGC AGC GGG CAG CAA      1346
Ser Asn His Ala Gly Gly Ile Leu Gly Gly Ile Ser Ser Gly Gln Gln
270                 275                 280                 285

ATC ATT GCC CAT ATG GCG CTG AAA CCG ACC TCC AGC ATT ACC GTG CCG      1394
Ile Ile Ala His Met Ala Leu Lys Pro Thr Ser Ser Ile Thr Val Pro
                290                 295                 300

GGT CGT ACC ATT AAC CGC TTT GGC GAA GAA GTT GAG ATG ATC ACC AAA      1442
Gly Arg Thr Ile Asn Arg Phe Gly Glu Glu Val Glu Met Ile Thr Lys
            305                 310                 315

GGC CGT CAC GAT CCC TGT GTC GGG ATC CGC GCA GTG CCG ATC GCA GAA      1490
Gly Arg His Asp Pro Cys Val Gly Ile Arg Ala Val Pro Ile Ala Glu
        320                 325                 330

GCG AAT GCT GGC GAT CGT TTT AAT GGA TCA CCT GTT ACG GCA ACG GGC      1538
Ala Asn Ala Gly Asp Arg Phe Asn Gly Ser Pro Val Thr Ala Thr Gly
    335                 340                 345

GCA AAA TGC CGA TGT GAA GAC TGA TATTCCACGC TGGTAAAAAA TGAATAAAAC      1592
Ala Lys Cys Arg Cys Glu Asp  *
350                 355

CGCGATTGCG CTGCTGGCTC TGCTTGCCAG TAGCGCCAGC CTGGCAGCGA CGCCGTGGCA    1652

AAAAATAACC CAACCTGTGC CGGGTAGCGC CAAATCGA                            1690
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 356 amino acids
(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Met Ala Gly Asn Thr Ile Gly Gln Leu Phe Arg Val Thr Thr Phe Gly
 1               5                  10                  15

Glu Ser His Gly Leu Ala Leu Gly Cys Ile Val Asp Gly Val Pro Pro
            20                  25                  30

Gly Ile Pro Leu Thr Glu Ala Asp Leu Gln His Asp Leu Asp Arg Arg
        35                  40                  45

Arg Pro Gly Thr Ser Arg Tyr Thr Thr Gln Arg Arg Glu Pro Asp Gln
    50                  55                  60

Val Lys Ile Leu Ser Gly Val Phe Glu Gly Val Thr Thr Gly Thr Ser
65                  70                  75                  80

Ile Gly Leu Leu Ile Glu Asn Thr Asp Gln Arg Ser Gln Asp Tyr Ser
                85                  90                  95

Ala Ile Lys Asp Val Phe Arg Pro Gly His Ala Asp Tyr Thr Tyr Glu
            100                 105                 110

Gln Lys Tyr Gly Leu Arg Asp Tyr Arg Gly Gly Gly Arg Ser Ser Ala
        115                 120                 125

Arg Glu Thr Ala Met Arg Val Ala Ala Gly Ala Ile Ala Lys Lys Tyr
    130                 135                 140

Leu Ala Glu Lys Phe Gly Ile Glu Ile Arg Gly Cys Leu Thr Gln Met
145                 150                 155                 160

Gly Asp Ile Pro Leu Asp Ile Lys Asp Trp Ser Gln Val Glu Gln Asn
                165                 170                 175

Pro Phe Phe Cys Pro Asp Pro Asp Lys Ile Asp Ala Leu Asp Glu Leu
            180                 185                 190

Met Arg Ala Leu Lys Lys Glu Gly Asp Ser Ile Gly Ala Lys Val Thr
        195                 200                 205

Val Val Ala Ser Gly Val Pro Ala Gly Leu Gly Glu Pro Val Phe Asp
    210                 215                 220

Arg Leu Asp Ala Asp Ile Ala His Ala Leu Met Ser Ile Asn Ala Val
225                 230                 235                 240

Lys Gly Val Glu Ile Gly Asp Gly Phe Asp Val Val Ala Leu Arg Gly
                245                 250                 255

Ser Gln Asn Arg Asp Glu Ile Thr Lys Asp Gly Phe Gln Ser Asn His
            260                 265                 270
```

```
Ala Gly Gly Ile Leu Gly Gly Ile Ser Ser Gly Gln Gln Ile Ile Ala
        275                 280                 285

His Met Ala Leu Lys Pro Thr Ser Ser Ile Thr Val Pro Gly Arg Thr
    290                 295                 300

Ile Asn Arg Phe Gly Glu Glu Val Glu Met Ile Thr Lys Gly Arg His
305                 310                 315                 320

Asp Pro Cys Val Gly Ile Arg Ala Val Pro Ile Ala Glu Ala Asn Ala
                325                 330                 335

Gly Asp Arg Phe Asn Gly Ser Pro Val Thr Ala Thr Gly Ala Lys Cys
            340                 345                 350

Arg Cys Glu Asp  *
        355
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1713 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: ompC of E.coli

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:491..1594

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
GTTAACAAGC GTTATAGTTT TTCTGTGGTA GCACAGAATA ATGAAAAGTG TGTAAAGAAG      60

GGTAAAAAAA ACCGAATGCG AGGCATCCGG TTGAAATAGG GGTAAACAGA CATTCAGAAA     120

TGAATGACGG TAATAAATAA AGTTAATGAT GATAGCGGGA GTTATTCTAG TTGCGAGTGA     180

AGGTTTTGTT TTGACATTCA GTGCTGTCAA ATACTTAAGA ATAAGTTATT GATTTTAACC     240

TTGAATTATT ATTGCTTGAT GTTAGGTGCT TATTTCGCCA TTCCGCAATA ATCTTAAAAA     300

GTTCCCTTGC ATTTACATTT TGAAACATCT ATAGCGATAA ATGAAACATC TTAAAAGTTT     360

TAGTATCATA TTCGTGTTGG ATTATTCTGC ATTTTTGGGG AGAATGGACT TGCCGACTGA     420

TTAATGAGGG TTAATCAGTA TGCAGTGGCA TAAAAAAGCA AATAAAGGCA TATAACAGAG     480
```

```
GGTTAATAAC ATG AAA GTT AAA GTA CTG TCC CTC CTG GTC CCA GCT CTG          529
           Met Lys Val Lys Val Leu Ser Leu Leu Val Pro Ala Leu
                   360                 365                 370

CTG GTA GCA GGC GCA GCA AAC GCT GCT GAA GTT TAC AAC AAA GAC GGC          577
Leu Val Ala Gly Ala Ala Asn Ala Ala Glu Val Tyr Asn Lys Asp Gly
               375                 380                 385

AAC AAA TTA GAT CTG TAC GGT AAA GTA GAC GGC CTG CAC TAT TTC TCT          625
Asn Lys Leu Asp Leu Tyr Gly Lys Val Asp Gly Leu His Tyr Phe Ser
           390                 395                 400

GAC AAC AAA GAT GTA GAT GGC GAC CAG ACC TAC ATG CGT CTT GGC TTC          673
Asp Asn Lys Asp Val Asp Gly Asp Gln Thr Tyr Met Arg Leu Gly Phe
       405                 410                 415

AAA GGT GAA ACT CAG GTT ACT GAC CAG CTG ACC GGT TAC GGC CAG TGG          721
Lys Gly Glu Thr Gln Val Thr Asp Gln Leu Thr Gly Tyr Gly Gln Trp
   420                 425                 430

GAA TAT CAG ATC CAG GGC AAC AGC GCT GAA AAC GAA AAC AAC TCC TGG          769
Glu Tyr Gln Ile Gln Gly Asn Ser Ala Glu Asn Glu Asn Asn Ser Trp
435                 440                 445                 450

ACC CGT GTG GCA TTC GCA GGT CTG AAA TTC CAG GAT GTG GGT TCT TTC          817
Thr Arg Val Ala Phe Ala Gly Leu Lys Phe Gln Asp Val Gly Ser Phe
               455                 460                 465

GAC TAC GGT CGT AAC TAC GGC GTT GTT TAT GAC GTA ACT TCC TGG ACC          865
Asp Tyr Gly Arg Asn Tyr Gly Val Val Tyr Asp Val Thr Ser Trp Thr
           470                 475                 480

GAC GTA CTG CCA GAA TTC GGT GGT GAC ACC TAC GGT TCT GAC AAC TTC          913
Asp Val Leu Pro Glu Phe Gly Gly Asp Thr Tyr Gly Ser Asp Asn Phe
       485                 490                 495

ATG CAG CAG CGT GGT AAC GGC TTC GCG ACC TAC CGT AAC ACT GAC TTC          961
Met Gln Gln Arg Gly Asn Gly Phe Ala Thr Tyr Arg Asn Thr Asp Phe
   500                 505                 510

TTC GGT CTG GTT GAC GGC CTG AAC TTT GCT GTT CAG TAC CAG GGT AAA         1009
Phe Gly Leu Val Asp Gly Leu Asn Phe Ala Val Gln Tyr Gln Gly Lys
515                 520                 525                 530

AAC GGC AAC CCA TCT GGT GAA GGC TTT ACT AGT GGC GTA ACT AAC AAC         1057
Asn Gly Asn Pro Ser Gly Glu Gly Phe Thr Ser Gly Val Thr Asn Asn
               535                 540                 545

GGT CGT GAC GCA CTG CGT CAA AAC GGC GAC GGC GTC GGC GGT TCT ATC         1105
Gly Arg Asp Ala Leu Arg Gln Asn Gly Asp Gly Val Gly Gly Ser Ile
           550                 555                 560

ACT TAT GAT TAC GAA GGT TTC GGT ATC GGT GGT GCG ATC TCC AGC TCC         1153
```

```
Thr Tyr Asp Tyr Glu Gly Phe Gly Ile Gly Gly Ala Ile Ser Ser Ser
       565                 570                 575

AAA CGT ACT GAT GCT CAG AAC ACC GCT GCT TAC ATC GGT AAC GGC GAC      1201
Lys Arg Thr Asp Ala Gln Asn Thr Ala Ala Tyr Ile Gly Asn Gly Asp
    580                 585                 590

CGT GCT GAA ACC TAC ACT GGT GGT CTG AAA TAC GAC GCT AAC AAC ATC      1249
Arg Ala Glu Thr Tyr Thr Gly Gly Leu Lys Tyr Asp Ala Asn Asn Ile
595                 600                 605                 610

TAC CTG GCT GCT CAG TAC ACC CAG ACC TAC AAC GCA ACT CGC GTA GGT      1297
Tyr Leu Ala Ala Gln Tyr Thr Gln Thr Tyr Asn Ala Thr Arg Val Gly
                615                 620                 625

TCC CTG GGT TGG GCG AAC AAA GCA CAG AAC TTC GAA GCT GTT GCT CAG      1345
Ser Leu Gly Trp Ala Asn Lys Ala Gln Asn Phe Glu Ala Val Ala Gln
            630                 635                 640

TAC CAG TTC GAC TTC GGT CTG CGT CCG TCC CTG GCT TAC CTG CAG TCT      1393
Tyr Gln Phe Asp Phe Gly Leu Arg Pro Ser Leu Ala Tyr Leu Gln Ser
       645                 650                 655

AAA GGT AAA AAC CTG GGT CGT GGC TAC GAC GAC GAA GAT ATC CTG AAA      1441
Lys Gly Lys Asn Leu Gly Arg Gly Tyr Asp Asp Glu Asp Ile Leu Lys
    660                 665                 670

TAT GTT GAT GTT GGT GCT ACC TAC TAC TTC AAC AAA AAC ATG TCC ACC      1489
Tyr Val Asp Val Gly Ala Thr Tyr Tyr Phe Asn Lys Asn Met Ser Thr
675                 680                 685                 690

TAC GTT GAC TAC AAA ATC AAC CTG CTG GAC GAC AAC CAG TTC ACT CGT      1537
Tyr Val Asp Tyr Lys Ile Asn Leu Leu Asp Asp Asn Gln Phe Thr Arg
                695                 700                 705

GAC GCT GGC ATC AAC ACT GAT AAC ATC GTA GCT CTG GGT CTG GTT TAC      1585
Asp Ala Gly Ile Asn Thr Asp Asn Ile Val Ala Leu Gly Leu Val Tyr
            710                 715                 720

CAG TTC TAA TCTCGATTGA TATCGAACAA GGGCCTGCGG GCCCTTTTTT              1634
Gln Phe  *
       725

CATTGTTTTC AGCGTACAAA CTCAGTTTTT TGGTGTACTC TTGCGACCGT TCGCATGAGG    1694

ATAATCACGT ACGGAAATA                                                 1713
```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 367 amino acids
(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```
Met Lys Val Lys Val Leu Ser Leu Leu Val Pro Ala Leu Leu Val Ala
 1               5                  10                  15

Gly Ala Ala Asn Ala Ala Glu Val Tyr Asn Lys Asp Gly Asn Lys Leu
            20                  25                  30

Asp Leu Tyr Gly Lys Val Asp Gly Leu His Tyr Phe Ser Asp Asn Lys
        35                  40                  45

Asp Val Asp Gly Asp Gln Thr Tyr Met Arg Leu Gly Phe Lys Gly Glu
    50                  55                  60

Thr Gln Val Thr Asp Gln Leu Thr Gly Tyr Gly Gln Trp Glu Tyr Gln
65                  70                  75                  80

Ile Gln Gly Asn Ser Ala Glu Asn Glu Asn Asn Ser Trp Thr Arg Val
                85                  90                  95

Ala Phe Ala Gly Leu Lys Phe Gln Asp Val Gly Ser Phe Asp Tyr Gly
            100                 105                 110

Arg Asn Tyr Gly Val Val Tyr Asp Val Thr Ser Trp Thr Asp Val Leu
        115                 120                 125

Pro Glu Phe Gly Gly Asp Thr Tyr Gly Ser Asp Asn Phe Met Gln Gln
    130                 135                 140

Arg Gly Asn Gly Phe Ala Thr Tyr Arg Asn Thr Asp Phe Phe Gly Leu
145                 150                 155                 160

Val Asp Gly Leu Asn Phe Ala Val Gln Tyr Gln Gly Lys Asn Gly Asn
                165                 170                 175

Pro Ser Gly Glu Gly Phe Thr Ser Gly Val Thr Asn Asn Gly Arg Asp
            180                 185                 190

Ala Leu Arg Gln Asn Gly Asp Gly Val Gly Gly Ser Ile Thr Tyr Asp
        195                 200                 205

Tyr Glu Gly Phe Gly Ile Gly Gly Ala Ile Ser Ser Ser Lys Arg Thr
    210                 215                 220

Asp Ala Gln Asn Thr Ala Ala Tyr Ile Gly Asn Gly Asp Arg Ala Glu
225                 230                 235                 240

Thr Tyr Thr Gly Gly Leu Lys Tyr Asp Ala Asn Asn Ile Tyr Leu Ala
                245                 250                 255

Ala Gln Tyr Thr Gln Thr Tyr Asn Ala Thr Arg Val Gly Ser Leu Gly
            260                 265                 270
```

```
Trp Ala Asn Lys Ala Gln Asn Phe Glu Ala Val Ala Gln Tyr Gln Phe
        275                 280                 285

Asp Phe Gly Leu Arg Pro Ser Leu Ala Tyr Leu Gln Ser Lys Gly Lys
    290                 295                 300

Asn Leu Gly Arg Gly Tyr Asp Asp Glu Asp Ile Leu Lys Tyr Val Asp
305                 310                 315                 320

Val Gly Ala Thr Tyr Tyr Phe Asn Lys Asn Met Ser Thr Tyr Val Asp
                325                 330                 335

Tyr Lys Ile Asn Leu Leu Asp Asp Asn Gln Phe Thr Arg Asp Ala Gly
            340                 345                 350

Ile Asn Thr Asp Asn Ile Val Ala Leu Gly Leu Val Tyr Gln Phe  *
        355                 360                 365
```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1808 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: ompF of E.coli

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:457..1545

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
AAAACTAATC CGCATTCTTA TTGCGGATTA GTTTTTTCTT AGCTAATAGC ACAATTTTCA      60

TACTATTTTT TGGCATTCTG GATGTCTGAA AGAAGATTTT GTGCCAGGTC GATAAAGTTT     120

CCATCAGAAA CAAAATTTCC GTTTAGTTAA TTTAAATATA AGGAAATCAT ATAAATAGAT     180

TAAAATTGCT GTAAATATCA TCACGTCTCT ATGGAAATAT GACGGTGTTC ACAAAGTTCC     240

TTAAATTTTA CTTTTGGTTA CATATTTTTT CTTTTTGAAA CCAAATCTTT ATCTTTGTAG     300

CACTTTCACG GTAGCGAAAC GTTAGTTTGA ATGGAAAGAT GCCTGCAGAC ACATAAAGAC     360

ACCAAACTCT CATCAATAGT TCCGTAAATT TTTATTGACA GAACTTATTG ACGGCAGTGG     420

CAGGTGTCAT AAAAAAAACC ATGAGGGTAA TAAATA ATG ATG AAG CGC AAT ATT       474
```

```
                                                    Met Met Lys Arg Asn Ile
                                                      1               5

CTG GCA GTG ATC GTC CCT GCT CTG TTA GTA GCA GGT ACT GCA AAC GCT          522
Leu Ala Val Ile Val Pro Ala Leu Leu Val Ala Gly Thr Ala Asn Ala
            10                  15                  20

GCA GAA ATC TAT AAC AAA GAT GGC AAC AAA GTA GAT CTG TAC GGT AAA          570
Ala Glu Ile Tyr Asn Lys Asp Gly Asn Lys Val Asp Leu Tyr Gly Lys
        25                  30                  35

GCT GTT GGT CTG CAT TAT TTT TCC AAG GGT AAC GGT GAA AAC AGT TAC          618
Ala Val Gly Leu His Tyr Phe Ser Lys Gly Asn Gly Glu Asn Ser Tyr
    40                  45                  50

GGT GGC AAT GGC GAC ATG ACC TAT GCC CGT CTT GGT TTT AAA GGG GAA          666
Gly Gly Asn Gly Asp Met Thr Tyr Ala Arg Leu Gly Phe Lys Gly Glu
55                  60                  65                  70

ACT CAA ATC AAT TCC GAT CTG ACC GGT TAT GGT CAG TGG GAA TAT AAC          714
Thr Gln Ile Asn Ser Asp Leu Thr Gly Tyr Gly Gln Trp Glu Tyr Asn
                75                  80                  85

TTC CAG GGT AAC AAC TCT GAA GGC GCT GAC GCT CAA ACT GGT AAC AAA          762
Phe Gln Gly Asn Asn Ser Glu Gly Ala Asp Ala Gln Thr Gly Asn Lys
            90                  95                  100

ACG CGT CTG GCA TTC GCG GGT CTT AAA TAC GCT GAC GTT GGT TCT TTC          810
Thr Arg Leu Ala Phe Ala Gly Leu Lys Tyr Ala Asp Val Gly Ser Phe
        105                 110                 115

GAT TAC GGC CGT AAC TAC GGT GTG GTT TAT GAT GCA CTG GGT TAC ACC          858
Asp Tyr Gly Arg Asn Tyr Gly Val Val Tyr Asp Ala Leu Gly Tyr Thr
    120                 125                 130

GAT ATG CTG CCA GAA TTT GGT GGT GAT ACT GCA TAC AGC GAT GAC TTC          906
Asp Met Leu Pro Glu Phe Gly Gly Asp Thr Ala Tyr Ser Asp Asp Phe
135                 140                 145                 150

TTC GTT GGT CGT GTT GGC GGC GTT GCT ACC TAT CGT AAC TCC AAC TTC          954
Phe Val Gly Arg Val Gly Gly Val Ala Thr Tyr Arg Asn Ser Asn Phe
                155                 160                 165

TTT GGT CTG GTT GAT GGC CTG AAC TTC GCT GTT CAG TAC CTG GGT AAA         1002
Phe Gly Leu Val Asp Gly Leu Asn Phe Ala Val Gln Tyr Leu Gly Lys
            170                 175                 180

AAC GAG CGT GAC ACT GCA CGC CGT TCT AAC GGC GAC GGT GTT GGC GGT         1050
Asn Glu Arg Asp Thr Ala Arg Arg Ser Asn Gly Asp Gly Val Gly Gly
        185                 190                 195

TCT ATC AGC TAC GAA TAC GAA GGC TTT GGT ATC GTT GGT GCT TAT GGT         1098
Ser Ile Ser Tyr Glu Tyr Glu Gly Phe Gly Ile Val Gly Ala Tyr Gly
    200                 205                 210
```

```
GCA GCT GAC CGT ACC AAC CTG CAA GAA GCT CAA CCT CTT GGC AAC GGT      1146
Ala Ala Asp Arg Thr Asn Leu Gln Glu Ala Gln Pro Leu Gly Asn Gly
215                 220                 225                 230

AAA AAA GCT GAA CAG TGG GCT ACT GGT CTG AAG TAC GAC GCG AAC AAC      1194
Lys Lys Ala Glu Gln Trp Ala Thr Gly Leu Lys Tyr Asp Ala Asn Asn
                235                 240                 245

ATC TAC CTG GCA GCG AAC TAC GGT GAA ACC CGT AAC GCT ACG CCG ATC      1242
Ile Tyr Leu Ala Ala Asn Tyr Gly Glu Thr Arg Asn Ala Thr Pro Ile
            250                 255                 260

ACT AAT AAA TTT ACA AAC ACC AGC GGC TTC GCC AAC AAA ACG CAA GAC      1290
Thr Asn Lys Phe Thr Asn Thr Ser Gly Phe Ala Asn Lys Thr Gln Asp
        265                 270                 275

GTT CTG TTA GTT GCG CAA TAC CAG TTC GAT TTC GGT CTG CGT CCG TCC      1338
Val Leu Leu Val Ala Gln Tyr Gln Phe Asp Phe Gly Leu Arg Pro Ser
    280                 285                 290

ATC GCT TAC ACC AAA TCT AAA GCG AAA GAC GTA GAA GGT ATC GGT GAT      1386
Ile Ala Tyr Thr Lys Ser Lys Ala Lys Asp Val Glu Gly Ile Gly Asp
295                 300                 305                 310

GTT GAT CTG GTG AAC TAC TTT GAA GTG GGC GCA ACC TAC TAC TTC AAC      1434
Val Asp Leu Val Asn Tyr Phe Glu Val Gly Ala Thr Tyr Tyr Phe Asn
                315                 320                 325

AAA AAC ATG TCC ACC TAT GTT GAC TAC ATC ATC AAC CAG ATC GAT TCT      1482
Lys Asn Met Ser Thr Tyr Val Asp Tyr Ile Ile Asn Gln Ile Asp Ser
            330                 335                 340

GAC AAC AAA CTG GGC GTA GGT TCA GAC GAC ACC GTT GCT GTG GGT ATC      1530
Asp Asn Lys Leu Gly Val Gly Ser Asp Asp Thr Val Ala Val Gly Ile
        345                 350                 355

GTT TAC CAG TTC TAA TAGCACACCT CTTTGTTAAA TGCCGAAAAA ACAGGACTTT      1585
Val Tyr Gln Phe  *
    360

GGTCCTGTTT TTTTTATACC TTCCAGAGCA ATCTCACGTC TTGCAAAAAC AGCCTGCGTT    1645

TTCATCAGTA ATAGTTGGAA TTTTGTAAAT CTCCCGTTAC CCTGATAGCG GACTTCCCTT    1705

CTGTAACCAT AATGGAACCT CGTCATGTTT GAGAACATTA CCGCCGCTCC TGCCGACCCG    1765

ATTCTGGGCC TGGCCGATCT GTTTCGTGCC GATGAACGTC CG                       1808
```

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 362 amino acids
(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
Met Met Lys Arg Asn Ile Leu Ala Val Ile Val Pro Ala Leu Leu Val
 1               5                  10                  15

Ala Gly Thr Ala Asn Ala Ala Glu Ile Tyr Asn Lys Asp Gly Asn Lys
            20                  25                  30

Val Asp Leu Tyr Gly Lys Ala Val Gly Leu His Tyr Phe Ser Lys Gly
        35                  40                  45

Asn Gly Glu Asn Ser Tyr Gly Gly Asn Gly Asp Met Thr Tyr Ala Arg
    50                  55                  60

Leu Gly Phe Lys Gly Glu Thr Gln Ile Asn Ser Asp Leu Thr Gly Tyr
65                  70                  75                  80

Gly Gln Trp Glu Tyr Asn Phe Gln Gly Asn Asn Ser Glu Gly Ala Asp
                85                  90                  95

Ala Gln Thr Gly Asn Lys Thr Arg Leu Ala Phe Ala Gly Leu Lys Tyr
            100                 105                 110

Ala Asp Val Gly Ser Phe Asp Tyr Gly Arg Asn Tyr Gly Val Val Tyr
        115                 120                 125

Asp Ala Leu Gly Tyr Thr Asp Met Leu Pro Glu Phe Gly Gly Asp Thr
    130                 135                 140

Ala Tyr Ser Asp Asp Phe Phe Val Gly Arg Val Gly Gly Val Ala Thr
145                 150                 155                 160

Tyr Arg Asn Ser Asn Phe Phe Gly Leu Val Asp Gly Leu Asn Phe Ala
                165                 170                 175

Val Gln Tyr Leu Gly Lys Asn Glu Arg Asp Thr Ala Arg Arg Ser Asn
            180                 185                 190

Gly Asp Gly Val Gly Gly Ser Ile Ser Tyr Glu Tyr Glu Gly Phe Gly
        195                 200                 205

Ile Val Gly Ala Tyr Gly Ala Ala Asp Arg Thr Asn Leu Gln Glu Ala
    210                 215                 220

Gln Pro Leu Gly Asn Gly Lys Lys Ala Glu Gln Trp Ala Thr Gly Leu
225                 230                 235                 240

Lys Tyr Asp Ala Asn Asn Ile Tyr Leu Ala Ala Asn Tyr Gly Glu Thr
                245                 250                 255

Arg Asn Ala Thr Pro Ile Thr Asn Lys Phe Thr Asn Thr Ser Gly Phe
            260                 265                 270
```

```
Ala Asn Lys Thr Gln Asp Val Leu Leu Val Ala Gln Tyr Gln Phe Asp
       275                 280                 285

Phe Gly Leu Arg Pro Ser Ile Ala Tyr Thr Lys Ser Lys Ala Lys Asp
    290                 295                 300

Val Glu Gly Ile Gly Asp Val Asp Leu Val Asn Tyr Phe Glu Val Gly
305                 310                 315                 320

Ala Thr Tyr Tyr Phe Asn Lys Asn Met Ser Thr Tyr Val Asp Tyr Ile
                325                 330                 335

Ile Asn Gln Ile Asp Ser Asp Asn Lys Leu Gly Val Gly Ser Asp Asp
            340                 345                 350

Thr Val Ala Val Gly Ile Val Tyr Gln Phe  *
       355                 360
```

CLAIMS

1. A bacterium attenuated by a non-reverting mutation in each of the *aroC* gene, the *ompF* gene and the *ompC* gene.

2. A bacterium according to claim 1 which infects by the oral route.

3. A bacterium according to claim 1 or 2 which is from the genera *Escherichia, Salmonella, Vibrio, Haemophilus, Neisseria, Yersinia, Bordetella* or *Brucella*.

4. A bacterium according to claim 3 which is a strain of *Escherichia coli, Salmonella typhimurium, Salmonella typhi, Salmonella enteritidis, Salmonella choleraesuis, Salmonella dublin, Haemophilus influenzae, Neisseria gonorrhoeae, Yersinia enterocolitica, Bordetella pertussis* or *Brucella abortus*.

5. A bacterium according to claim 4 which is a strain of enterotoxigenic *E.coli* (ETEC.

6. A bacterium according to any one of the preceding

claims which is further attenuted by a mutation in a fourth gene.

7. A bacterium according to claim 6 wherein the fourth gene is *aroA, aroE, pur, htrA, galE, cya, crp, phoP* or *surA*.

8. A bacterium according to any one of the preceding claims, wherein the mutation in each gene is a defined mutation.

9. A bacterium according to any one of the preceding claims, wherein the mutation in each gene is deletion of the entire coding sequence.

10. A bacterium according to any one of the preceding claims which has been genetically engineered to express a heterologous antigen.

11. A bacterium according to claim 10, wherein expression of the antigen is driven by the *nirB* promoter or the *htrA* promoter.

12. A vaccine comprising a bacterium as defined in any one of the preceding claims and a pharmaceutically acceptable carrier or diluent.

13. A bacterium as defined in any one of claims 1 to 11 for use in a method of vaccinating a human or animal.

14. Use of a bacterium as defined in any one of claims 1 to 11 for the manufacture of a medicament for vaccinating a human or animal.

15. A method of raising an immune response in a mammalian host, which comprises administering to the host a bacterium attenuated by a non-reverting mutation in each of the *aroC* gene, the *ompF* gene and the *ompC* gene.

ABSTRACT

ATTENUATED BACTERIA USEFUL IN VACCINES

The invention provides a bacterium attenuated by a non-reverting mutation in each of the aroC gene, the ompF gene and the ompC gene. The bacterium is useful as a vaccine. The bacterium may, for example, be an attenuated strain of E.coli useful in vaccination against diarrhoea.

Figure 1

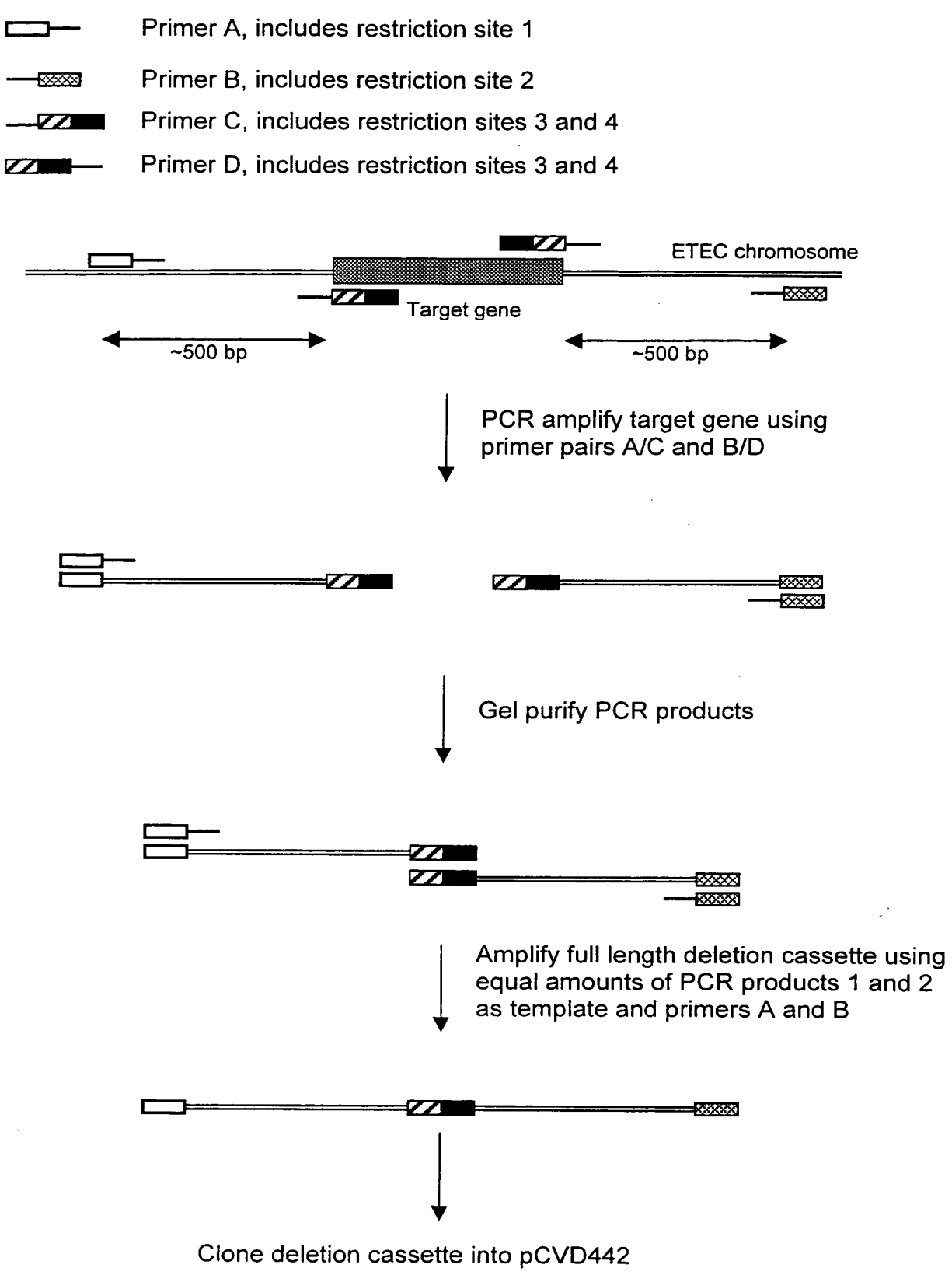

Primer A, includes restriction site 1
Primer B, includes restriction site 2
Primer C, includes restriction sites 3 and 4
Primer D, includes restriction sites 3 and 4
ETEC chromosome
Target gene
~500 bp
~500 bp
PCR amplify target gene using primer pairs A/C and B/D
Gel purify PCR products
Amplify full length deletion cassette using equal amounts of PCR products 1 and 2 as template and primers A and B
Clone deletion cassette into pCVD442

Figure 2

aroC
w.t. AAACACAACAATAACGGAGCGTG**ATG**---TAAAAATGAATAAAACCGCGATTG CG
deletion AAACACAACAATAACGGAGCC*CTCGAGGCATGC*TGAATAAAATGAATAAAACCGCGATTG CG

htrA
w.t TGTTAATCGAGAXTGAAATAC**ATG**AA---AG**TAA**TCTCCCTCAACCCCTTCCT GAA
deletion TGTTAATCGAGAXTGAAATAC*CTCGAGTCTAGA*CTCCCTCAACCCCTTCCT GAA

ompC
w.t. ATATAACAGAGGGTTAATAAC**ATG**AAA---CAGTTC**TAA** TCTCGATTGATATCGAAC
deletion ATATAACAGAGGGTTAATAAC*GCTAAGCCTCGAG***TAA** TCTCGATTGATATCGAAC

ompF
w.t. AAACCATGAGGGTAATAAAATA**ATG**ATGAAGCGC---CCAGTTC**TAA** TAGCACACCTCTTTGTTA
deletion AAACCATGAGGGTAATAAAATAga*GCTAAGCCTCGAG*CAGTTC**TAA** TAGCACACCTCTTTGTTA

ompR
w.t. CGAACCTTTGGGAGTACAAACA**ATG**CAA---AAGCA**TGA** GGCGATTGCGCTTCTCGCCA
deletion CGAACCTTTGGGAGTACAAACA*GCTAAGCGCATGC***GA** GGCGATTGCGCTTCTCGCCA

Bold – Stop and start codons
Italics – restriction enzyme sites introduced
Underlined – primer binding sites
Lower case – extra n.t added to primers to avoid primer dimer formation
--- wild type gene
N.B. *aroC* deletion removes 16 n.t. 3' to the stop codon

Figure 3

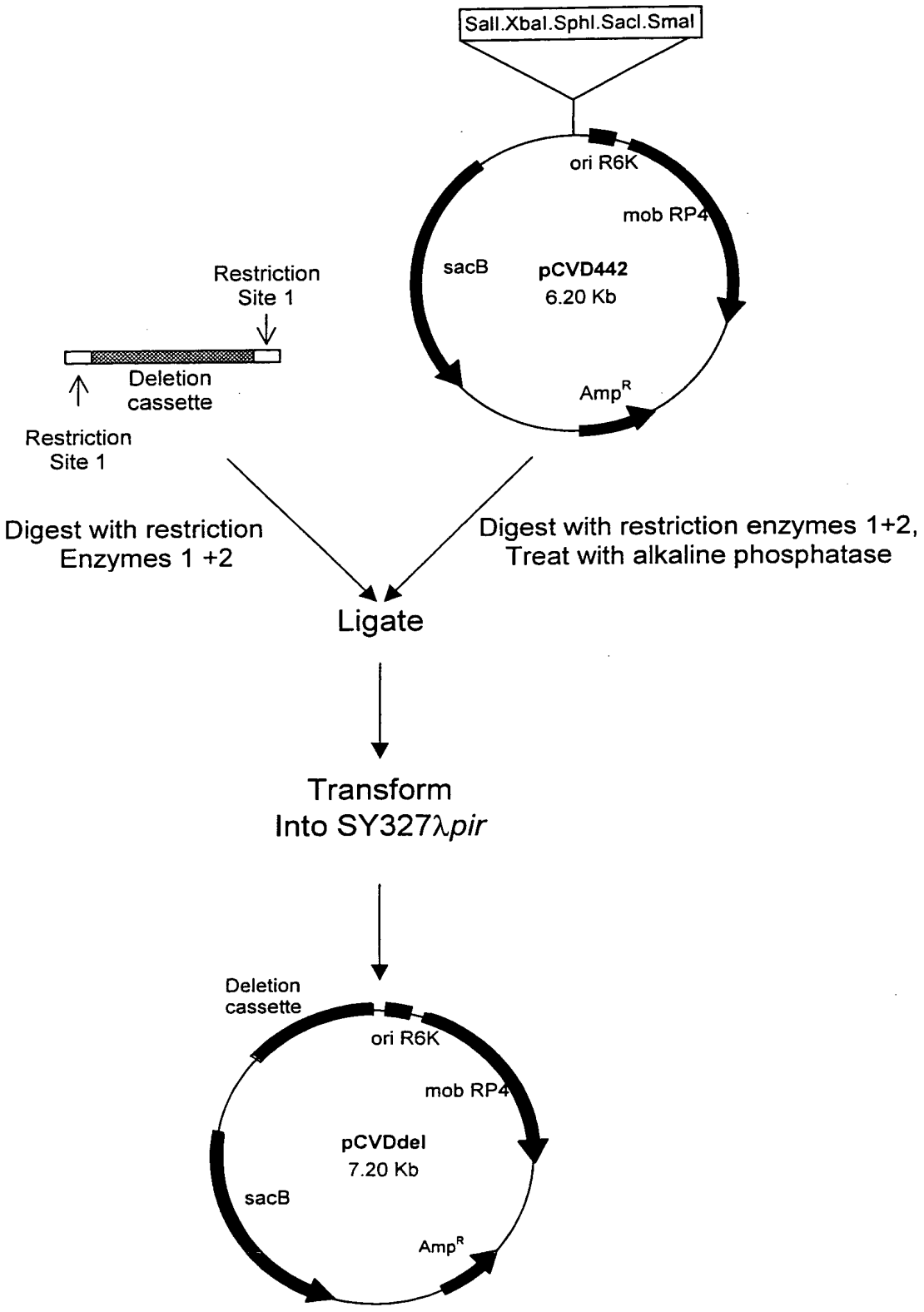

SalI.XbaI.SphI.SacI.SmaI
ori R6K
mob RP4
sacB
pCVD442
6.20 Kb
Amp^R
Restriction
Site 1
Deletion
cassette
Restriction
Site 1
Digest with restriction
Enzymes 1 +2
Digest with restriction enzymes 1+2,
Treat with alkaline phosphatase
Ligate
Transform
Into SY327λpir
Deletion
cassette
ori R6K
mob RP4
pCVDdel
7.20 Kb
sacB
Amp^R

Figure 4

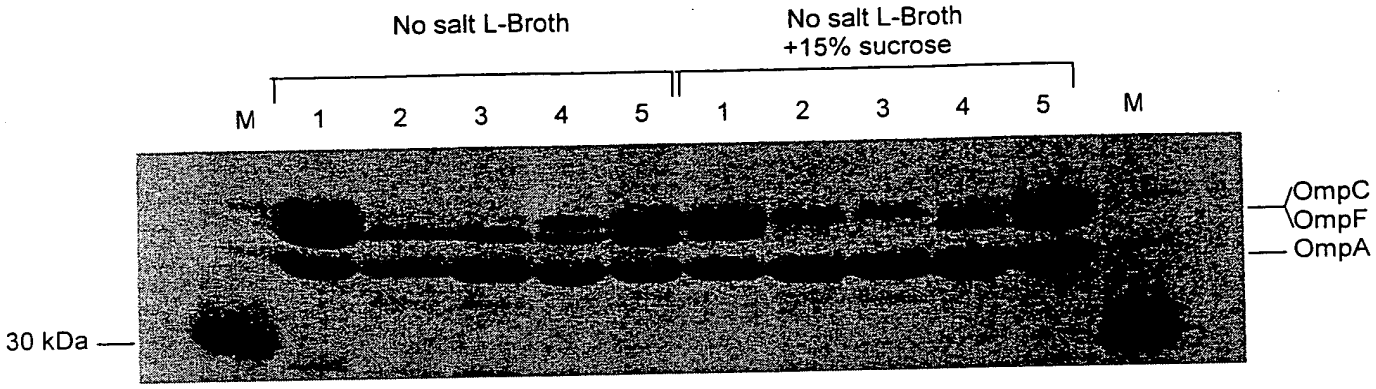

No salt L-Broth
No salt L-Broth
+15% sucrose
M
1
2
3
4
5
1
2
3
4
5
M
OmpC
OmpF
OmpA
30 kDa

Figure 5

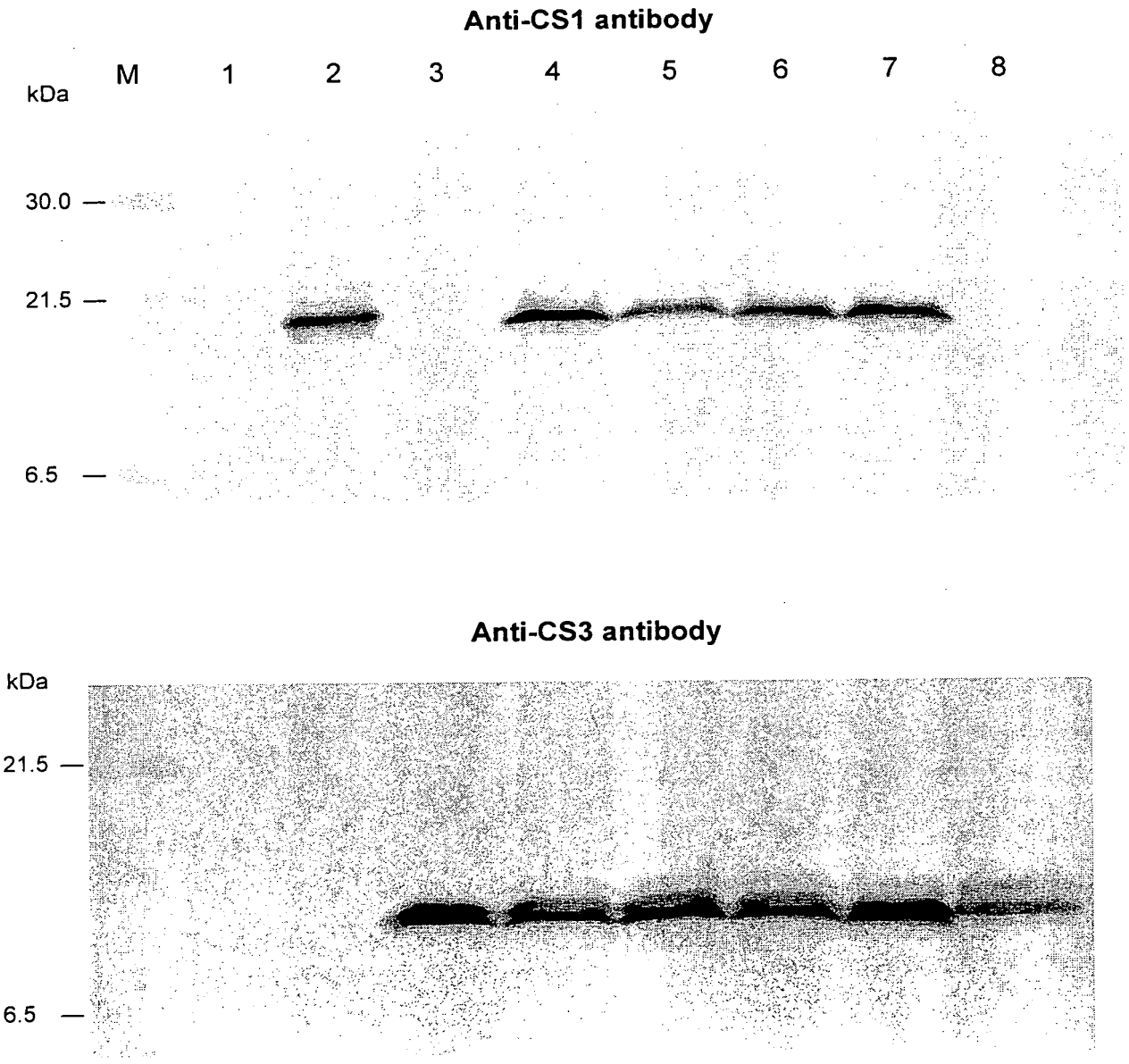
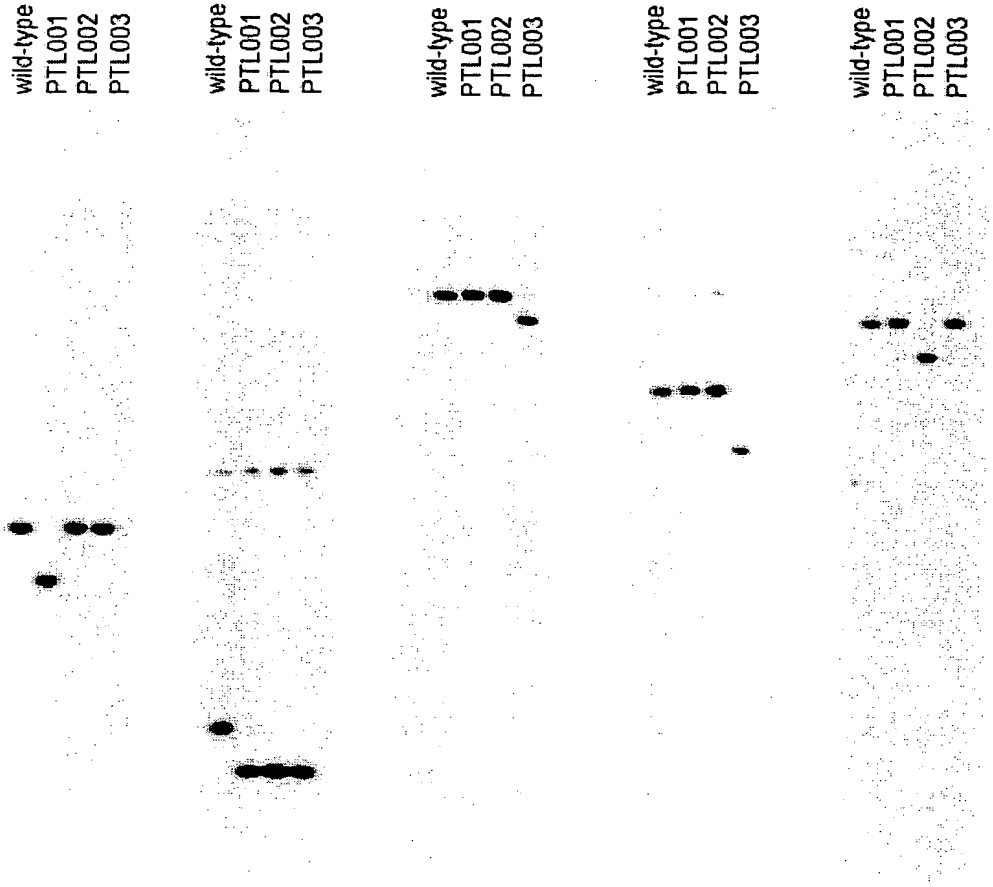

Anti-CS1 antibody
M
1
2
3
4
5
6
7
8
kDa
30.0
21.5
6.5
Anti-CS3 antibody
kDa
21.5
6.5

Figure 6


wild-type
PTL001
PTL002
PTL003
wild-type
PTL001
PTL002
PTL003
wild-type
PTL001
PTL002
PTL003
wild-type
PTL001
PTL002
PTL003
wild-type
PTL001
PTL002
PTL003

PG GB 9900935
JA Kemp a
14/4/99